

# SENSING TENSORS WITH GAUSSIAN FILTERS

STÉPHANE CHRÉTIEN AND TIANWEN WEI

**ABSTRACT.** Sparse recovery from linear Gaussian measurements has been the subject of much investigation since the breakthrough papers [6] and [11] on Compressed Sensing. Application to sparse vectors and sparse matrices via least squares penalized with sparsity promoting norms is now well understood using tools such as Gaussian mean width, statistical dimension and the notion of descent cones [22] [23]. Extension of these ideas to low rank tensor recovery is starting to enjoy considerable interest due to its many potential applications to Independent Component Analysis, Hidden Markov Models and Gaussian Mixture Models [1], hyperspectral image analysis [25], to name a few. In this paper, we demonstrate that the recent approach of [23] provides very useful error bounds in the tensor setting using the nuclear norm or the Romera-Paredes–Pontil [20] penalization.

## 1. INTRODUCTION

Real tensors, i.e. multidimensional arrays of real numbers, have been recently a subject of great interest in the applied mathematics community. We refer to [15] and [13] for modern references on this subject. It has become quite clear nowadays that real symmetric tensors such as cumulants up to fourth order play a very important role in many applications in statistics, machine learning and signal processing; see for instance [14] for a general survey. Research on applications of tensors has been increasing in the recent years with very important conceptual contributions such as proposed in [3], [2], [9], [4] and the very nice survey of applications in [1]. In particular, certain Gaussian Mixture Models (GMM) can be estimated using this approach. The same is also true for Independent Component Analysis (ICA) and Hidden Markov Models (HMM). Nonsymmetric tensors also occur frequently in applications as 3D images such as in medical imaging and hyperspectral image processing [14], [24].

In some applications, the tensor is observed through the operation of random filtering, e.g. taking the scalar product with an i.i.d. Gaussian random vector. Gaussian random sensing has been thoroughly investigated in recent years and can be recast as sparse recovery for one dimensional tensors (i.e. vectors), low rank recovery for bidimensional tensors (i.e. matrices). See e.g. [23] for a tutorial on this topic.

The goal of this short note is to show that the results of [23] can easily be extended to tensors. First, we consider nuclear norm minimization. Next, we consider recovery via minimization of the Romero-Paredes–Pontil functional.

## 2. MAIN FACTS ABOUT TENSORS

Let  $D$  and  $n_1, \dots, n_D$  be positive integers. Let  $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}$  denote a  $D$ -dimensional array of real numbers. We will also denote such arrays as tensors.

**2.1. Basic notations and operations.** A subtensor of  $\mathcal{X}$  is a tensor obtained by fixing some of its coordinates. As an example, fixing one coordinate  $i_d = k$  in  $\mathcal{X}$  for some  $k \in \{1, \dots, n_d\}$

yields a tensor in  $\mathbb{R}^{n_1 \times \dots \times n_{d-1} \times n_{d+1} \times \dots \times n_D}$ . In the sequel, we will denote this subtensor of  $\mathcal{X}$  by  $\mathcal{X}_{i_d=k}$ .

The fibers of a tensor are particular subtensors that have only one mode, i.e. obtained by fixing every coordinate except one. The mode- $d$  fibers are the vectors

$$(\mathcal{X}_{i_1, \dots, i_{d-1}, i_d, i_{d+1}, \dots, i_D})_{i_d=1, \dots, n_d}.$$

They extend the notion of columns and rows from the matrix to the tensor framework. For a matrix, the mode-1 fibers are the columns and the mode-2 fibers are the rows.

The mode- $d$  matricization  $\mathcal{X}_{(d)}$  of  $\mathcal{X}$  is obtained by forming the matrix whose rows are the mode- $d$  fibers of the tensor, arranged in an cyclic ordering; see [14] for details.

The mode- $d$  multiplication of a tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}$  by a matrix  $U \in \mathbb{R}^{n'_d \times n_d}$ , denoted by  $\mathcal{X} \times_d U$ , gives a tensor in  $\mathbb{R}^{n_1 \times \dots \times n'_d \times \dots \times n_D}$ . It is defined as

$$(\mathcal{X} \times_d U)_{i_1, \dots, i_{d-1}, i'_d, i_{d+1}, \dots, i_D} = \sum_{i_d=1}^{n_d} \mathcal{X}_{i_1, \dots, i_{d-1}, i_d, i_{d+1}, \dots, i_D} U_{i'_d, i_d}.$$

Last, we denote by  $\|\cdot\|_F$  the Frobenius norm, i.e.:

$$\|\mathcal{X}\|_F = \left( \sum_{i_1=1}^{n_1} \dots \sum_{i_D=1}^{n_D} \mathcal{X}_{i_1, \dots, i_D}^2 \right)^{1/2}, \quad \forall \mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}.$$

**2.2. Higher Order Singular Value Decomposition (HOSVD).** The Tucker decomposition of a tensor is a very useful decomposition. It can be chosen so that after appropriate orthogonal transformations, one can reveal a tensor  $\mathcal{S}$  hidden inside  $\mathcal{X}$  enjoying interesting rank and orthogonality properties. In this contribution, we will make use of the HOSVD, a generalization of the matrix SVD to the tensor setting based on Tucker decomposition.

**Theorem 2.1** (HOSVD [10]). *Every tensor  $\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}$  can be written as*

$$(2.1) \quad \mathcal{X} = \mathcal{S}(\mathcal{X}) \times_1 U^{(1)} \times_2 U^{(2)} \dots \times_D U^{(D)},$$

where each  $U^{(d)} \in \mathbb{R}^{n_d \times n_d}$  is an orthogonal matrix and  $\mathcal{S}(\mathcal{X}) \in \mathbb{R}^{n_1 \times \dots \times n_D}$  is a tensor of the same size as  $\mathcal{X}$  with the following properties:

1. For all possible values of  $d, \alpha$  and  $\beta$  subject to  $\alpha \neq \beta$ , subtensors  $\mathcal{S}(\mathcal{X})_{i_d=\alpha}$  and  $\mathcal{S}(\mathcal{X})_{i_d=\beta}$  are orthogonal, i.e.

$$\langle \mathcal{S}(\mathcal{X})_{i_d=\alpha}, \mathcal{S}(\mathcal{X})_{i_d=\beta} \rangle = 0.$$

2. For all possible values of  $d$ , there holds :

$$\|\mathcal{S}(\mathcal{X})_{i_d=1}\|_F \geq \|\mathcal{S}(\mathcal{X})_{i_d=2}\|_F \geq \dots \geq \|\mathcal{S}(\mathcal{X})_{i_d=n_d}\|_F \geq 0.$$

3. The quantities  $\|\mathcal{S}(\mathcal{X})_{i_d=k}\|_F$  for  $k = 1, \dots, n_d$  are the singular values of the mode- $d$  matricization  $\mathcal{X}_{(d)}$  of  $\mathcal{X}$  and the columns of  $U^{(d)}$  are the corresponding singular vectors.

Let  $\otimes$  denote the standard Kronecker product for matrices. Then it follows from (2.1) that

$$(2.2) \quad \mathcal{X}_{(d)} = U^{(d)} \mathcal{S}_{(d)}(\mathcal{X}) \left( U^{(d+1)} \otimes \dots \otimes U^{(D)} \otimes U^{(1)} \otimes \dots \otimes U^{(d-1)} \right)^t,$$

where  $\mathcal{S}_{(d)}(\mathcal{X})$  is the mode- $d$  matricization of  $\mathcal{S}(\mathcal{X})$ . Taking the (usual) SVD of the matrix  $\mathcal{X}_{(d)}$

$$\mathcal{X}_{(d)} = U^{(d)} \Sigma^{(d)} V^{(d)t}$$

and based on (2.2), we get

$$\mathcal{S}_{(d)}(\mathcal{X}) = \Sigma^{(d)} V^{(d)t} \left( U^{(d+1)} \otimes \dots \otimes U^{(D)} \otimes U^{(1)} \otimes \dots \otimes U^{(d-1)} \right).$$

**2.3. The spectrum.** The mode- $d$  spectrum is defined as the vector of singular values of  $\mathcal{X}_{(d)}$  and we will denote it by  $\sigma^{(d)}(\mathcal{X})$ . Notice that this construction implies that  $\mathcal{S}(\mathcal{X})$  has orthonormal fibers for every modes. With a slight abuse of notation, we will denote by  $\sigma$  the mapping which to each tensor  $\mathcal{X}$  assigns the vector  $1/\sqrt{D} (\sigma^{(1)}, \dots, \sigma^{(D)})$  of all mode- $d$  singular spectra.

**2.4. Tensor norms.** We can define several tensor norms on  $\mathbb{R}^{n_1 \times \dots \times n_D}$ . The first one is a natural extension of the Frobenius norm or Hilbert-Schmidt norm from matrices to tensors. We start by defining the following scalar product on  $\mathbb{R}^{n_1 \times \dots \times n_D}$ :

$$\langle \mathcal{X}, \mathcal{Y} \rangle \stackrel{\text{def}}{=} \sum_{i_1=1}^{n_1} \dots \sum_{i_D=1}^{n_D} x_{i_1, \dots, i_D} y_{i_1, \dots, i_D}.$$

Using this scalar product, we can also define the Frobenius norm as

$$\|\mathcal{X}\|_F \stackrel{\text{def}}{=} \sqrt{\langle \mathcal{X}, \mathcal{X} \rangle}.$$

One may also define an "operator norm" in the same manner as for matrices as follows

$$\|\mathcal{X}\| \stackrel{\text{def}}{=} \max_{\substack{u^{(d)} \in \mathbb{R}^{n_d}, \|u^{(d)}\|_2 = 1 \\ d=1, \dots, D}} \langle \mathcal{X}, u^{(1)} \otimes \dots \otimes u^{(D)} \rangle$$

We also define the

$$(2.3) \quad \|\mathcal{X}\|_* \stackrel{\text{def}}{=} \max_{\substack{\mathcal{Y} \in \mathbb{R}^{n_1 \times \dots \times n_D} \\ \|\mathcal{Y}\| \leq 1}} \langle \mathcal{X}, \mathcal{Y} \rangle.$$

The norm  $\|\cdot\|_*$  can be interpreted as a generalization of the nuclear norm for matrices. They can be shown to be equal for certain class of tensors such as Orthogonally Decomposable tensors [7]. Another interesting function is the Romera-Paredes-Pontil functional.

### 3. TENSOR RECOVERY BASED ON RANDOM MEASUREMENTS USING CONVEX OPTIMIZATION

**3.1. Previous works.** Our goal is to estimate an unknown but structured  $n_1 \times \dots \times n_D$  tensor  $\mathcal{X}^\sharp$  from  $m$  linear observations given by

$$y_i = \langle \mathcal{G}_i, \mathcal{X}^\sharp \rangle, \quad i = 1, \dots, m.$$

The unknown tensor of interests  $\mathcal{X}^\sharp$  although resides in a extremely high dimensional data space, it has a low-rank structure in many applications. The general problem of estimating a low rank tensor has applications in many different areas, both theoretical and applied. We refer the readers to [2, 19].

For a tensor of low Tucker rank, the matrix unfolding along each mode has low rank. Given observations  $y_1, \dots, y_m$ , we would like to attempt to recover  $\mathcal{X}^\sharp$  by minimizing some combination of the ranks of the unfoldings, over all tensors  $\mathcal{X}^\sharp$  that are consistent with our observations. This yields the following optimization problem:

$$(3.4) \quad \min_{\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}} \sum_{d=1}^d \text{rank}(\mathcal{X}_{(d)}) \quad \text{s.t.} \quad \langle \mathcal{G}_i, \mathcal{X} \rangle = y_i, \quad i = 1, \dots, m.$$

Optimization problem (3.4), although intuitive, is non-convex and NP hard. A natural convex surrogate of (3.4) can be obtained by replacing the rank with matrix nuclear norms [12]. The resulting optimization problem becomes

$$(3.5) \quad \min_{\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}} \sum_{d=1}^d \|\mathcal{X}_{(d)}\|_* \quad \text{s.t.} \quad \langle \mathcal{G}_i, \mathcal{X} \rangle = y_i, \quad i = 1, \dots, m.$$

This optimization problem was first introduced by [18, 21] and has been used successfully in a number of applications [17].

In this work, we are going to consider a related but somewhat different optimization problem:

$$(3.6) \quad \min_{\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}} \|\mathcal{X}\|_* \quad \text{s.t.} \quad \langle \mathcal{G}_i, \mathcal{X} \rangle = y_i, \quad i = 1, \dots, m,$$

where the tensor nuclear norm  $\|\cdot\|_*$  is defined in (2.3). We point out that in general  $\|\mathcal{X}\|_* \neq \frac{1}{D} \sum_{d=1}^D \|\mathcal{X}_{(d)}\|_*$  but the two quantities coincide when  $\mathcal{X}$  is an orthogonally decomposable tensor. The main reason to consider (3.6) is that the method established in [23] for Gaussian random filter can be easily generalized to the tensor framework.

### 3.2. Recovery by nuclear norm minimization.

**Theorem 3.1.** *Assume that  $\mathcal{G}_i, i = 1, \dots, m$  are independent  $n_1 \times \dots \times n_D$  random Gaussian tensors with independent  $\mathcal{N}(0, 1)$  entries. Let  $\hat{\mathcal{X}}$  to be a solution of the convex program*

$$\min_{\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}} \|\mathcal{X}\|_* \quad \text{s.t.} \quad \langle \mathcal{G}_i, \mathcal{X} \rangle = y_i, \quad i = 1, \dots, m.$$

Then

$$(3.7) \quad \mathbb{E} \left\| \hat{\mathcal{X}} - \mathcal{X}^\# \right\|_F \leq 4\sqrt{\pi} \frac{\sqrt{n_1} + \dots + \sqrt{n_D}}{m} \|\mathcal{X}^\#\|_*.$$

*Proof.* We consider the following set of  $n_1 \times \dots \times n_D$  tensors:

$$K \stackrel{\text{def}}{=} \{\mathcal{X} : \|\mathcal{X}\|_* \leq \|\mathcal{X}^\#\|_*\}.$$

Applying Theorem 6.2 from [23], we obtain

$$\mathbb{E} \left[ \sup_{\substack{\mathcal{X} \in K \\ \langle \mathcal{G}_i, \mathcal{X} \rangle = y_i, i=1, \dots, m}} \|\hat{\mathcal{X}} - \mathcal{X}\|_F \right] \leq \sqrt{2\pi} \cdot \frac{w(K)}{\sqrt{m}},$$

where  $w(K)$  denotes the Gaussian mean width [23] of set  $K$ . By the symmetry of  $K$ , we have

$$w(K) = \mathbb{E} \left[ \sup_{\mathcal{X} \in K-K} \langle \mathcal{G}, \mathcal{X} \rangle \right] = 2 \mathbb{E} \sup_{\mathcal{X} \in K} \langle \mathcal{G}, \mathcal{X} \rangle,$$

where  $\mathcal{G}$  is a Gaussian random tensors with  $\mathcal{N}(0, 1)$  entries. Then using the inequality  $\langle \mathcal{G}, \mathcal{X} \rangle \leq \|\mathcal{G}\| \cdot \|\mathcal{X}\|_*$  and Lemma 4.1, we obtain

$$w(K) \leq 2 \mathbb{E} \left[ \sup_{\mathcal{X} \in K} \|\mathcal{G}\| \cdot \|\mathcal{X}\|_* \right] \leq 2 (\sqrt{n_1} + \dots + \sqrt{n_D}) \|\mathcal{X}^\#\|_*.$$

Then bound (3.7) follows.  $\square$

**3.3. The Romera-Paredes–Pontil relaxation.** In the sequel, let us denote by  $N = \sum_{i=1}^D n_i$ . The Romera-Paredes–Pontil function on  $\mathbb{R}^N$ , denoted by  $\omega_\alpha^{**}$ , is the convex envelope of the cardinality function  $\|\cdot\|_0$  on the  $\ell_2$ -ball of radius  $\alpha$ . Its conjugate  $\omega_\alpha^*$  is defined as

$$(3.8) \quad \omega_\alpha^*(g) = \sup_{\|s\|_2 \leq \alpha} \langle g, s \rangle - \|s\|_0.$$

By conjugate duality, we have for all  $g, s$  in  $\mathbb{R}^N$

$$(3.9) \quad \langle g, s \rangle \leq \omega_\alpha^*(g) + \omega_\alpha^{**}(s).$$

Taking  $g$  such that  $\omega_\alpha^*(g) = 1$  and  $\omega_\alpha^{**}(s) = 1$ , we have

$$\langle g, s \rangle \leq 2.$$

Therefore, for any  $g$  and  $s$ , we have

$$(3.10) \quad \langle g, s \rangle \leq 2 \omega_\alpha^*(g) \omega_\alpha^{**}(s).$$

Moreover, by the Von Neumann's trace inequality for tensors [8], we have

$$(3.11) \quad \langle \mathcal{G}, \mathcal{X} \rangle \leq \langle \sigma(\mathcal{G}), \sigma(\mathcal{X}) \rangle$$

combining (3.10) and (3.11), we obtain that

$$(3.12) \quad \langle \mathcal{G}, \mathcal{X} \rangle \leq 2 \omega_\alpha^*(\sigma(\mathcal{G})) \omega_\alpha^{**}(\sigma(\mathcal{X})).$$

based on these results, we obtain the following theorem.

**Theorem 3.2.** Assume that  $\mathcal{G}_i$ ,  $i, \dots, m$  are independent  $n_1 \times \dots \times n_D$  random Gaussian tensors with independent  $\mathcal{N}(0, 1)$  entries. Let  $\hat{\mathcal{X}}$  be a solution of the convex program

$$(3.13) \quad \min_{\mathcal{X} \in \mathbb{R}^{n_1 \times \dots \times n_D}} \omega_\alpha^{**}(\sigma(\mathcal{X})) \quad \text{s.t.} \quad \langle \mathcal{G}_i, \mathcal{X} \rangle = y_i, \quad i = 1, \dots, m.$$

Then

$$(3.14) \quad \mathbb{E} \left[ \|\hat{\mathcal{X}} - \mathcal{X}^\# \|_F \right] \leq \frac{64 \alpha \sqrt{\pi}}{\sqrt{m}} \left[ \left( \sqrt{\log(p)} \frac{(1+N)^{D+1} - 2^{D+1}}{D+1} + \frac{1}{4} \sqrt{\pi q} \left( N + \frac{N(1+N)^D - 2^D}{D} \right. \right. \right. \\ \left. \left. \left. + \frac{2^{D+1} - (1+N)^{D+1}}{D(D+1)} \right) \right) - \frac{N(1+N)}{2} \right] \omega_\alpha^{**}(\sigma(\mathcal{X}^\#)),$$

where  $n^* = \max\{n_1, \dots, n_D\}$ ,  $p = C_{D-1}^{r-1} 3^{rn^*}$  and  $q = rn_* + (r/D)^D$ .

*Proof.* We consider the following set of  $n_1 \times \dots \times n_D$  tensors:

$$K \stackrel{\text{def}}{=} \{ \mathcal{X} : \omega_\alpha^{**}(\sigma(\mathcal{X})) \leq \omega_\alpha^{**}(\sigma(\mathcal{X}^\#)) \}.$$

Exactly as for the proof of Theorem 3.1, applying Theorem 6.2 from [23], we obtain

$$\mathbb{E} \left[ \sup_{\substack{\mathcal{X} \in K \\ \langle \mathcal{G}_i, \mathcal{X} \rangle = y_i, i=1, \dots, n}} \|\hat{\mathcal{X}} - \mathcal{X}\|_F \right] \leq \sqrt{2\pi} \cdot \frac{w(K)}{\sqrt{m}}.$$

and using the obvious symmetry of  $K$ , we have

$$w(K) = \mathbb{E} \left[ \sup_{\mathcal{X} \in K-K} \langle \mathcal{G}, \mathcal{X} \rangle \right] = 2 \mathbb{E} \left[ \sup_{\mathcal{X} \in K} \langle \mathcal{G}, \mathcal{X} \rangle \right],$$

where  $\mathcal{G}$  is a Gaussian random tensors with  $\mathcal{N}(0, 1)$  entries. Then using inequality (3.12), we obtain

$$w(K) \leq 4 \mathbb{E} \left[ \omega_{\alpha}^*(\sigma(\mathcal{G})) \right] \omega_{\alpha}^{**}(\sigma(\mathcal{X}))$$

and using Lemma 4.3, we get

$$w(K) \leq 32\sqrt{2} \alpha \left( \left( \sqrt{\log(p)} \frac{(1+N)^{D+1} - 2^{D+1}}{D+1} + \frac{1}{4} \sqrt{\pi q} \left( N + \frac{N(1+N)^D - 2^D}{D} \right. \right. \right. \\ \left. \left. \left. + \frac{2^{D+1} - (1+N)^{D+1}}{D(D+1)} \right) \right) - \frac{N(1+N)}{2} \right) \omega_{\alpha}^{**}(\sigma(\mathcal{X}^{\sharp})).$$

Then bound (3.14) follows.  $\square$

#### 4. SOME RESULTS ON GAUSSIAN TENSORS

##### 4.1. The spectral norm of a Gaussian tensor.

**Lemma 4.1.** *Let*

$$\mathcal{X} = (\mathcal{X}_{i_1, \dots, i_D})_{i_1=1, \dots, n_1, \dots, i_D=1, \dots, n_D}.$$

*be a tensor with i.i.d. standard Gaussian entries. Then*

$$\mathbb{E} [\|\mathcal{X}\|] \leq \sum_{i=1}^D \sqrt{n_i}.$$

*Proof.* Consider the following stochastic process indexed by  $(u^{(1)}, \dots, u^{(D)})$ , where  $u^{(d)} \in \mathcal{S}^{n_d-1}$  for  $d = 1, \dots, D$

$$X_{u^{(1)}, \dots, u^{(D)}} \stackrel{\text{def}}{=} \langle \mathcal{X}, u^{(1)} \otimes \dots \otimes u^{(D)} \rangle.$$

Let us compute the variance of the increments of this random process. For any  $u^{(d)}, v^{(d)} \in \mathcal{S}^{n_d-1}$  for  $d = 1, \dots, D$ , we have

$$\mathbb{E} \left[ \left\| X_{u^{(1)}, \dots, u^{(D)}} - X_{v^{(1)}, \dots, v^{(D)}} \right\|_F^2 \right] \\ = \mathbb{E} \left[ \sum_{i_1, \dots, i_D=1}^{n_1, \dots, n_D} \mathcal{X}_{i_1, \dots, i_D}^2 \left( \prod_{d=1}^D u_{i_d}^{(d)} - \prod_{d=1}^D v_{i_d}^{(d)} \right)^2 \right]$$

and since the entries of  $\mathcal{X}$  have unit variance, we get

$$\mathbb{E} \left[ \left\| X_{u^{(1)}, \dots, u^{(D)}} - X_{v^{(1)}, \dots, v^{(D)}} \right\|_F^2 \right] \\ = \sum_{i_1, \dots, i_D=1}^{n_1, \dots, n_D} \left( \prod_{d=1}^D u_{i_d}^{(d)} - \prod_{d=1}^D v_{i_d}^{(d)} \right)^2 \\ = \left\| u^{(1)} \otimes \dots \otimes u^{(D)} - v^{(1)} \otimes \dots \otimes v^{(D)} \right\|_F^2$$

Moreover, since

$$\left\| u^{(1)} \otimes \dots \otimes u^{(D)} - v^{(1)} \otimes \dots \otimes v^{(D)} \right\|_F^2 \leq \sum_{i=1}^D \|u^{(i)} - v^{(i)}\|_F^2,$$

we obtain that

$$\mathbb{E} \left[ \left\| X_{u^{(1)}, \dots, u^{(D)}} - X_{v^{(1)}, \dots, v^{(D)}} \right\|_F^2 \right] = \sum_{i=1}^D \|u^{(i)} - v^{(i)}\|_F^2$$

Now let us consider another random process indexed by  $(u^{(1)}, \dots, u^{(D)})$ :

$$Y_{u^{(1)}, \dots, u^{(D)}} \stackrel{\text{def}}{=} \sum_{d=1}^D \langle g^{(d)}, u^{(d)} \rangle,$$

where  $g^{(d)} \sim \mathcal{N}(0, I_{n_d})$ ,  $d = 1, \dots, D$  are independent Gaussian random vectors. It is easy to see that the variance of the increments of  $Y$  is

$$\begin{aligned} \mathbb{E} \left\| Y_{u^{(1)}, \dots, u^{(D)}} - Y_{v^{(1)}, \dots, v^{(D)}} \right\|^2 &= \mathbb{E} \left\| \sum_{d=1}^D \langle g^{(d)}, u^{(d)} - v^{(d)} \rangle \right\|^2 \\ &= \sum_{d=1}^D \|u^{(d)} - v^{(d)}\|_F^2. \end{aligned}$$

Therefore, the variance of the increments of  $Y$  is greater than or equal to the variance of the increments of  $X$ . Therefore, we can apply Slepian's lemma [16] and obtain

$$\begin{aligned} \mathbb{E} [\|\mathcal{X}\|] &= \mathbb{E} \left[ \sup_{\substack{u^{(i)}: \|u^{(i)}\|=1 \\ i=1, \dots, D}} X(u^{(1)}, \dots, u^{(D)}) \right] \\ &\leq \mathbb{E} \left[ \sup_{\substack{u^{(i)}: \|u^{(i)}\|=1 \\ i=1, \dots, D}} Y(u^{(1)}, \dots, u^{(D)}) \right] \\ &= \mathbb{E} \left[ \sup_{\substack{u^{(i)}: \|u^{(i)}\|=1 \\ i=1, \dots, D}} \sum_{i=1}^D \langle g^{(i)}, u^{(i)} \rangle \right] \\ &= \mathbb{E} \left[ \sum_{i=1}^D \|g^{(i)}\| \right] \end{aligned}$$

and by Jensen's inequality

$$\mathbb{E} [\|\mathcal{X}\|] \leq \sum_{i=1}^D \sqrt{n_i}$$

where the last inequality is derived from Jensen's inequality.  $\square$

**4.2. The entropy of the set of tensors with Frobenius norm equal to  $\alpha$  and given Tucker rank.** Define the set

$$\mathbb{T}_{r,\alpha} = \left\{ \mathcal{W} \in \mathbb{R}^{n_1 \times \dots \times n_D} \mid \|\mathcal{W}\|_F \leq \alpha, \quad \sum_{d=1}^D \text{rank}(\mathcal{W}^{(d)}) = r \right\}.$$

Then, we have that

**Theorem 4.2.** *For each  $\varepsilon$ , the set  $\mathbb{T}_{r,\alpha}$  has an  $\varepsilon'$ -net of size  $N(\varepsilon)$  with*

$$N(\varepsilon', r) \leq C_{D-1}^{r-1} \left( \frac{3}{\alpha} \right)^{rn_*} \left( \frac{3\alpha}{\varepsilon} \right)^{rn_* + (r/D)^D}$$

where  $n_* = \max\{n_1, \dots, n_D\}$  and

$$\varepsilon' = \varepsilon + \alpha \left( (1 + r\varepsilon)^D - 1 \right).$$

*Proof.* For any  $n$  and  $\nu$  in  $\mathbb{N}$  with  $n \geq \nu$ , let  $\mathfrak{D}_{n,\nu}$  be the Stiefel manifold defined by

$$\mathfrak{D}_{n,\nu} = \{U \in \mathbb{R}^{n \times \nu} \mid U^t U = I_\nu\}.$$

Then, it was proved in [5, Lemma 3.1] that there exists an  $\varepsilon$ -covering number of size  $(9/\varepsilon)^{n\nu}$ . Let  $\mathcal{P}(r, D)$  denote the set of integer partitions of  $r$  using no more than  $D$  integers. Now for each  $\nu = (\nu_1, \dots, \nu_D) \in \mathcal{P}(r, D)$ , define the set

$$\mathfrak{L}(\nu) = \{\mathcal{S} \in \mathbb{R}^{\nu_1 \times \dots \times \nu_D} \mid \|\mathcal{S}\|_F = \alpha\}.$$

It is well known [16] that the unit sphere of  $\mathbb{R}^m$  admits an  $\varepsilon$ -net of size less than  $(3/\varepsilon)^m$ . Using this fact, we easily obtain an  $\varepsilon$ -net of  $\mathfrak{L}(r)$  of size no larger than  $(3\alpha/\varepsilon)^{\nu_1 \times \dots \times \nu_D}$ .

Next, we are going to determine the size of  $\varepsilon$ -net covering the set

$$\mathfrak{T}(\nu) = \left\{ \mathcal{S} \times_1 U^{(1)} \times_2 \dots \times_D U^{(D)} \mid \mathcal{S} \in \mathfrak{L}(\nu), U^{(d)} \in \mathfrak{D}_{n_d, \nu_d}, d = 1, \dots, D \right\}$$

Denote  $\Delta\mathcal{S} = \mathcal{S} - \mathcal{S}_0$  and  $\Delta U^{(d)} = U^{(d)} - U_0^{(d)}$  for  $d = 1, \dots, D$ . Then

$$\begin{aligned} \mathcal{W} &= \mathcal{S} \times_1 U^{(1)} \times_2 \dots \times_D U^{(D)} \\ &= (\mathcal{S}_0 + \Delta\mathcal{S}) \times_1 (U_0^{(1)} + \Delta U^{(1)}) \times_2 \dots \times_D (U_0^{(D)} + \Delta U^{(D)}) \\ &= \mathcal{S}_0 \times_1 (U_0^{(1)} + \Delta U^{(1)}) \times_2 \dots \times_D (U_0^{(D)} + \Delta U^{(D)}) \\ &\quad + \Delta\mathcal{S} \times_1 (U_0^{(1)} + \Delta U^{(1)}) \times_2 \dots \times_D (U_0^{(D)} + \Delta U^{(D)}), \end{aligned}$$

in which

$$\begin{aligned} &\mathcal{S}_0 \times_1 (U_0^{(1)} + \Delta U^{(1)}) \times_2 \dots \times_D (U_0^{(D)} + \Delta U^{(D)}) \\ &= \mathcal{S}_0 \times_1 U_0^{(1)} \times_2 \dots \times_D U_0^{(D)} + \sum_{d=1}^D \left( \mathcal{S}_0 \times_d \Delta U^{(d)} \prod_{\substack{i=1 \\ i \neq d}}^D \times_i U_0^{(i)} \right) + \dots \\ &\quad + \sum_{\substack{d_1, \dots, d_k=1, 1 \leq k \leq D \\ d_p \neq d_q, \forall p \neq q}}^D \left( \mathcal{S}_0 \prod_{j=1}^k \times_{d_j} \Delta U^{(d_j)} \prod_{\substack{i=1 \\ i \neq d_1, \dots, d_k}}^D \times_i U_0^{(i)} \right) + \dots \\ &\quad + \mathcal{S}_0 \times_1 \Delta U_0^{(1)} \times_2 \dots \times_D \Delta U_0^{(D)}. \end{aligned}$$



Since

$$\begin{aligned} \left\| \Delta \mathcal{S} \times_1 (U_0^{(1)} + \Delta U^{(1)}) \times_2 \cdots \times_D (U_0^{(D)} + \Delta U^{(D)}) \right\|_F &= \|\Delta \mathcal{S}\|_F \\ \left\| \mathcal{S}_0 \prod_{j=1}^k \times_{d_j} \Delta U^{(d_j)} \prod_{\substack{i=1 \\ i \neq d_1, \dots, d_k}}^D \times_i U_0^{(i)} \right\|_F &\leq \alpha \varepsilon^k \prod_{j=1}^k \nu_{d_j} \end{aligned}$$

It follows that

$$\begin{aligned} & \left\| \mathcal{W} - \mathcal{S}_0 \times_1 U_0^{(1)} \times_2 \cdots \times_D U_0^{(D)} \right\|_F \\ & \leq \|\Delta \mathcal{S}\|_F + \sum_{d=1}^D \alpha \varepsilon \nu_d + \cdots + \sum_{\substack{d_1, \dots, d_k=1, 1 \leq k \leq D \\ d_p \neq d_q, \forall p \neq q}}^D \alpha \varepsilon^k \prod_{j=1}^k \nu_{d_j} + \cdots + \alpha \varepsilon^D \prod_{d=1}^D \nu_d \\ & \leq \varepsilon + \sum_{d=1}^D \alpha \varepsilon r + \cdots + \sum_{\substack{d_1, \dots, d_k=1, 1 \leq k \leq D \\ d_p \neq d_q, \forall p \neq q}}^D \alpha \varepsilon^k \left(\frac{r}{k}\right)^k + \cdots + \alpha \varepsilon^D \left(\frac{r}{D}\right)^D \\ & = \varepsilon + \alpha \sum_{k=1}^D C_D^k \varepsilon^k \left(\frac{r}{k}\right)^k \\ & \leq \varepsilon + \alpha \left( (1 + r\varepsilon)^D - 1 \right). \end{aligned}$$

Now let us rewrite the set  $\mathbb{T}$  as the following:

$$\begin{aligned} \mathbb{T}_{r, \alpha} &= \left\{ \mathcal{W} \in \mathbb{R}^{n_1 \times \cdots \times n_D} \mid \|\mathcal{W}\|_F \leq \alpha, \sum_{d=1}^D \text{rank}(\mathcal{W}^{(d)}) = r \right\} \\ &= \left\{ \mathcal{S} \times_1 U^{(1)} \times_2 \cdots \times_D U^{(D)} \mid \nu \in \mathcal{P}(r, D), \mathcal{S} \in \mathcal{L}(\nu), U^{(d)} \in \mathfrak{D}_{n_d, \nu_d}, d = 1, \dots, D \right\} \end{aligned}$$

and denote

$$\varepsilon' = \varepsilon + \alpha \left( (1 + r\varepsilon)^D - 1 \right).$$

Summing up our discussion, we conclude that there exists an  $\varepsilon$ -net of  $\mathbb{T}_{r, \alpha}$  with covering number

$$N(\varepsilon', r) \leq \sum_{\nu \in \mathcal{P}(r, D)} \left( \frac{9}{\varepsilon} \right)^{n_1 \nu_1 + \cdots + n_D \nu_D} \left( \frac{3\alpha}{\varepsilon} \right)^{\nu_1 \times \cdots \times \nu_D},$$

and since the cardinality of  $\mathcal{P}(r, D)$  equals  $C_{D-1}^{r-1}$ , we obtain that

$$\begin{aligned}
N(\varepsilon', r) &\leq C_{D-1}^{r-1} \max_{\nu \in \mathcal{P}(r, D)} \left( \frac{9}{\varepsilon} \right)^{n_1 \nu_1 + \dots + n_D \nu_D} \left( \frac{3\alpha}{\varepsilon} \right)^{\nu_1 \times \dots \times \nu_D} \\
&\leq C_{D-1}^{r-1} \max_{\nu \in \mathcal{P}(r, D)} \left( \frac{9}{\varepsilon} \right)^{n_1 \nu_1 + \dots + n_D \nu_D} \max_{\nu \in \mathcal{P}(r, D)} \left( \frac{3\alpha}{\varepsilon} \right)^{\nu_1 \times \dots \times \nu_D} \\
&\leq C_{D-1}^{r-1} \left( \frac{9}{\varepsilon} \right)^{rn_*} \left( \frac{3\alpha}{\varepsilon} \right)^{(r/D)^D} \\
&= C_{D-1}^{r-1} \left( \frac{3}{\alpha} \right)^{rn_*} \left( \frac{3\alpha}{\varepsilon} \right)^{rn_* + (r/D)^D},
\end{aligned}$$

where  $n_* = \max\{n_1, \dots, n_D\}$ . □

**4.3. The dual Romera-Paredes–Pontil function of Gaussian tensors.** In this section, we study the expected value of the evaluation at a Gaussian random tensor of the dual  $\omega_\alpha^*$  of the Romera-Paredes–Pontil function  $\omega_\alpha^{**}$ . We will need some further notations. For any vector  $s \in \mathbb{R}^N$ . Using (3.8), one easily obtains

$$(4.15) \quad \omega_\alpha^*(\mathcal{X}) \leq \sup_{\|w\|_2 \leq \alpha} \langle \sigma(\mathcal{X}), w \rangle - \|w\|_0.$$

We also have, by equation (7) in [20],

$$\omega_\alpha^*(\mathcal{X}) \leq \alpha \max_{r=0, \dots, N} \|\sigma_{1:r}^\perp(\mathcal{X})\|_2 - r.$$

We have the following result.

**Lemma 4.3.** *Let*

$$\mathcal{X} = (\mathcal{X}_{i_1, \dots, i_D})_{i_1=1, \dots, n_1, \dots, i_D=1, \dots, n_D}.$$

*be a tensor with i.i.d. standard Gaussian entries. Then*

$$\begin{aligned}
\mathbb{E}[\omega_\alpha^*(\mathcal{X})] &\leq 8\sqrt{2} \alpha \left( \sqrt{\log(p)} \frac{(1+N)^{D+1} - 2^{D+1}}{D+1} + \frac{1}{4} \sqrt{\pi q} \left( N + \frac{N(1+N)^D - 2^D}{D} \right. \right. \\
&\quad \left. \left. + \frac{2^{D+1} - (1+N)^{D+1}}{D(D+1)} \right) \right) - \frac{N(1+N)}{2}
\end{aligned}$$

*Proof.* Using (4.16), and the tensor Von Neumann inequality [8, Theorem 1], one obtains that

$$(4.16) \quad \omega_\alpha^*(\mathcal{X}) \leq \max_{r=1, \dots, N} \sup_{\substack{\|\mathcal{W}\|_F \leq \alpha \\ \sum_{d=1}^D \text{rank}(\mathcal{W}^{(d)}) = r}} \langle \mathcal{X}, \mathcal{W} \rangle - r$$

$$(4.17) \quad \leq \alpha \sum_{r=1, \dots, N} \sup_{\substack{\|\mathcal{W}\|_F \leq 1 \\ \sum_{d=1}^D \text{rank}(\mathcal{W}^{(d)}) = r}} \langle \mathcal{X}, \mathcal{W} \rangle - r$$

Since we must enforce the constraint  $\|\mathcal{W}\|_F \leq 1$ , Dudley's entropy bound says that [16]

$$(4.18) \quad \mathbb{E}[\omega_\alpha^*(\mathcal{X})] = 8\sqrt{2} \alpha \sum_{r=1, \dots, N} \int_0^1 \sqrt{\log(N(\varepsilon))} d\varepsilon - r.$$

Let us compute the integral term. We have

$$\begin{aligned}
& \int_0^1 \sqrt{\log(N(\varepsilon'))} d\varepsilon' \\
&= \int_0^1 \left[ \log \left( C_{D-1}^{r-1} (3)^{rn_*} \left( \frac{3}{\varepsilon} \right)^{rn_* + (r/D)^D} \right) \right]^{1/2} (1 + rD(1 + r\varepsilon)^{D-1}) d\varepsilon \\
&\leq \int_0^1 \sqrt{\log(p)} (1 + rD(1 + r\varepsilon)^{D-1}) d\varepsilon + \int_0^1 \sqrt{q \log(3/\varepsilon)} (1 + rD(1 + r\varepsilon)^{D-1}) d\varepsilon,
\end{aligned}$$

where  $p = C_{D-1}^{r-1} 3^{rn_*}$  and  $q = rn_* + (r/D)^D$ . We have

$$\begin{aligned}
& \int_0^1 \sqrt{\log(p)} (1 + rD(1 + r\varepsilon)^{D-1}) d\varepsilon \\
&= \sqrt{\log(p)} + rD\sqrt{\log(p)} \int_0^1 (1 + r\varepsilon)^{D-1} d\varepsilon \\
&= \sqrt{\log(p)} + rD\sqrt{\log(p)} \int_1^{r+1} z^{D-1} \frac{1}{r} dz \\
&= \sqrt{\log(p)} + \sqrt{\log(p)} ((r+1)^D - 1) \\
&= \sqrt{\log(p)} (r+1)^D,
\end{aligned}$$

and

$$\begin{aligned}
& \int_0^1 \sqrt{q \log(3/\varepsilon)} (1 + rD(1 + r\varepsilon)^{D-1}) d\varepsilon \\
&\leq \sqrt{q} (1 + rD(1 + r)^{D-1}) \int_0^1 \sqrt{\log(3/\varepsilon)} d\varepsilon \\
&= \sqrt{q} (1 + rD(1 + r)^{D-1}) \int_{\sqrt{\log 3}}^{+\infty} x^2 e^{-x^2} dx \\
&< \frac{1}{4} \sqrt{\pi q} (1 + rD(1 + r)^{D-1}).
\end{aligned}$$

Therefore

$$\begin{aligned}
& \int_0^1 \sqrt{\log(N(\varepsilon'))} d\varepsilon' \\
&\leq \sqrt{\log(p)} (1 + r)^D + \frac{1}{4} \sqrt{\pi q} (1 + rD(1 + r)^{D-1}),
\end{aligned}$$

where  $p = C_{D-1}^{r-1} (3)^{rn_*}$  and  $q = rn_* + (r/D)^D$ . Plugging this result into (4.18), we obtain

$$(4.19) \quad \mathbb{E} [\omega_\alpha^*(\mathcal{X})] \leq 8\sqrt{2} \alpha \sum_{r=1, \dots, N} \left( \sqrt{\log(p)} (1 + r)^D + \frac{1}{4} \sqrt{\pi q} (1 + rD(1 + r)^{D-1}) \right) - r.$$

Approximating sums by integrals, we thus obtain

$$\begin{aligned} \mathbb{E}[\omega_{\alpha}^*(\mathcal{X})] &\leq 8\sqrt{2}\alpha \left( \sqrt{\log(p)} \frac{(1+N)^{D+1} - 2^{D+1}}{D+1} + \frac{1}{4}\sqrt{\pi q} \left( N + \frac{N(1+N)^D - 2^D}{D} \right. \right. \\ &\quad \left. \left. + \frac{2^{D+1} - (1+N)^{D+1}}{D(D+1)} \right) \right) - \frac{N(1+N)}{2} \end{aligned}$$

as announced.  $\square$

## REFERENCES

1. Ge R.-Hsu D. Anandkumar, A., S. M. Kakade, and M. Telgarsky, *Tensor decompositions for learning latent variable models*, The Journal of Machine Learning Research **15** (2014), no. 1, 2773–2832.
2. Hsu D. Anandkumar, A. and S. M. Kakade, *A method of moments for mixture models and hidden markov models*, arXiv preprint arXiv:1203.0683 (2012).
3. Liu Y.-K. Hsu D. J. Foster D. P. Anandkumar, A. and S. M. Kakade, *A spectral algorithm for latent dirichlet allocation*, Advances in Neural Information Processing Systems, 2012, pp. 917–925.
4. J. Brachat, P. Comon, B. Mourrain, and E. Tsigaridas, *Symmetric tensor decomposition*, Linear Algebra and its Applications **433** (2010), no. 11, 1851–1872.
5. Emmanuel J Candes and Yaniv Plan, *Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements*, Information Theory, IEEE Transactions on **57** (2011), no. 4, 2342–2359.
6. Emmanuel J Candès, Justin Romberg, and Terence Tao, *Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information*, Information Theory, IEEE Transactions on **52** (2006), no. 2, 489–509.
7. S. Chrétien and T. Wei, *The subdifferential of some tensor norms*, (in preparation).
8. ———, *Von neumann’s trace inequality for tensors*, Linear Algebra and its Applications (to appear).
9. P. Comon, G. Golub, L.-H. Lim, and B. Mourrain, *Symmetric tensors and symmetric tensor rank*, SIAM Journal on Matrix Analysis and Applications **30** (2008), no. 3, 1254–1279.
10. Lieven De Lathauwer, Bart De Moor, and Joos Vandewalle, *A multilinear singular value decomposition*, SIAM journal on Matrix Analysis and Applications **21** (2000), no. 4, 1253–1278.
11. David L Donoho, *Compressed sensing*, Information Theory, IEEE Transactions on **52** (2006), no. 4, 1289–1306.
12. Maryam Fazel, *Matrix rank minimization with applications*, Elec Eng Dept Stanford University **54** (2002), 1–130.
13. W. Hackbusch, *Tensor spaces and numerical tensor calculus*, vol. 42, Springer Science & Business Media, 2012.
14. T. G. Kolda and B. W. Bader, *Tensor decompositions and applications*, SIAM review **51** (2009), no. 3, 455–500.
15. J.M. Landsberg, *Tensors: geometry and applications*, vol. 128, American Mathematical Soc., 2012.
16. Michel Ledoux and Michel Talagrand, *Probability in banach spaces: isoperimetry and processes*, vol. 23, Springer Science & Business Media, 2013.
17. Nan Li and Baoxin Li, *Tensor completion for on-board compression of hyperspectral images*, Image Processing (ICIP), 2010 17th IEEE International Conference on, Sept 2010, pp. 517–520.
18. Ji Liu, P. Musialski, P. Wonka, and Jieping Ye, *Tensor completion for estimating missing values in visual data*, Pattern Analysis and Machine Intelligence, IEEE Transactions on **35** (2013), no. 1, 208–220.
19. Bernardino Romera-paredes, Hane Aung, Nadia Bianchi-berthouze, and Massimiliano Pontil, *Multilinear multitask learning*, Proceedings of the 30th International Conference on Machine Learning (ICML-13) (Sanjoy Dasgupta and David Mcallester, eds.), vol. 28, JMLR Workshop and Conference Proceedings, May 2013, pp. 1444–1452.
20. Bernardino Romera-Paredes and Massimiliano Pontil, *A new convex relaxation for tensor completion*, Advances in Neural Information Processing Systems, 2013, pp. 2967–2975.

21. Marco Signoretto, Lieven De Lathauwer, and Johan A. K. Suykens<sup>†</sup>, *Nuclear Norms for Tensors and Their Use for Convex Multilinear Estimation*, Tech. report, ESAT-SISTA, K.U.Leuven (Leuven, Belgium), 2010.
22. Joel A Tropp, *Convex recovery of a structured signal from independent random linear measurements*, arXiv preprint arXiv:1405.1102 (2014).
23. R. Vershynin, *Estimation in high dimensions: a geometric perspective*, arXiv preprint arXiv:1405.5103 (2014).
24. Q. Zhang, H. Wang, R. J Plemmons, and V. Pauca, *Tensor methods for hyperspectral data analysis: a space object material identification study*, JOSA A **25** (2008), no. 12, 3001–3012.
25. Qiang Zhang, Han Wang, Robert J Plemmons, and V Pauca, *Tensor methods for hyperspectral data analysis: a space object material identification study*, JOSA A **25** (2008), no. 12, 3001–3012.

LABORATOIRE DE MATHÉMATIQUES, UMR 6623, UNIVERSITÉ DE FRANCHE-COMTÉ, 16 ROUTE DE GRAY,, 25030 BESANCON, FRANCE

*E-mail address:* `stephane.chretien,tianwen.wei@univ-fcomte.fr`